The journey of a team of engineers in learning packaging technology
1- Who we are, what we do?

- 4 engineers from 2 bioinformatics teams (INRA & IFREMER)
1- Who we are, what we do?

- 4 engineers, 2 distant places
- A lot of time installing and upgrading hundreds of tools
  - More or less easy to install
  - Pipelines: lot of dependencies

How to facilitate such installations for all bioinformatics teams?

- Let’s take a real example: FROGS
2- not the meal... but the metabarcoding analysis pipeline

FROGS: Find, Rapidly, OTUs with Galaxy Solution.


Bioinfo analysis:
- Reads Preprocessing
- Clustering
- Chimera removing
- OTUs affiliation
- Phylogenetic tree

Biostats analysis:
- Composition visualization
- Alpha diversity
- Beta diversity
- NMDS analysis
- Multivariate analysis
not us... but the metabarcodel analysis pipeline

- Presented at GCC 2016
- Nice for biologists
- Successful... on June 2018:
  - 20 trainings
  - 200 people trained
  - 17 citations ("FROGS", Escudie et al., 2017)
  - ~400 visit/month of the website frogs.toulouse.inra.fr
  - Deployed on many Galaxy instances
- Not so nice for engineers responsible for tools installation...
3- FROGS: installation process

Installation:
- From Github
- Manually
- Lot of dependencies
- Conflicts of versions

Source code

20 wrappers

github.com/geraldinepascal/FROGS

~20 dependencies

Manual installation

R-phyloseq
blast
vsearch
swarm
flash
cutadapt

...
3- FROGS accessibility

Accessibility for Users
✓

Accessibility for Admins
✗
(people in charge of tools installation)

How to automate the installation process?
4- FROGS: technologies available for software packaging

Dependencies & wrappers

Automatic installation

- Conda
- Docker
- Galaxy Toolshed
- Jenkins
- CircleCI
- OpenStack
- Maven
- Django
- Nginx
- Ansible
- Java Development Kit
- Spark
- Puppet Labs
- Node.js
- Vagrant
- Bamboo
4- FROGS: technologies available for software packaging

- Dependencies & wrappers
- Automatic installation

Galaxy Community Advices

- Conda
- Bioconda
- Docker
- Maven
- Python
- Ansible
- Node.js
- Nginx
- Jenkins
- OpenStack
- Java
- Spark
- Ansible
- Vagrant
- Bamboo
4- FROGS: technologies available for software packaging

- We don’t have any experience in packaging technologies
- Are these technologies easy to learn?
- Will we be able to package FROGS in a reasonable amount of time?

... so, let’s try them on FROGS!
5- A journey of a team of engineers in learning packaging technology

As a reminder: two teams in two different locations

Organize collaborative work sessions using:

- “appear.in” on the web to run online meetings
- a shared workspace on IFB cloud (Linux VM)
- “screen” to share simultaneously several terminals
  - text/code review
  - tools execution: conda packaging, planemo tests, ...
  - log control
- “vim” as our favourite user-friendly text/code editor

Organize regular meetings with FROGS dev-team (Géraldine and Maria)
5- A journey of a team of engineers in learning packaging technology

- **Step 1**: Follow the GTN materials for conda recipe
  
  (RTFM)

- **Step 2**: Check that all FROGS dependencies are available as conda (bio)-packages
  - Over more than 20 dependencies, the only missing one was r-phangorn
  - ... but sometimes you just have to wait a few days ⇒ what a dynamic community!
5- A journey of a team of engineers in learning packaging technology

- **Step 3:** Create a **conda env** with FROGS dependencies => ouch !!!
  - **Problem:** Conflict of dependencies versions between dev team and channels
  - **Solution:** Let **Conda** install compatible versions for dependencies
5- A journey of a team of engineers in learning packaging technology

- **Step 4:** Install FROGS manually and failed to run unit tests => ouch !!!
  - **Problem:** Dependencies installed by conda are not compatible regarding FROGS code

Conda took last versions

Incompatibilities
5- A journey of a team of engineers in learning packaging technology

- **Step 4:** Install FROGS manually and failed to run unit tests => ouch !!!
  - **Problem:** Dependencies installed by conda are *not compatible* regarding FROGS code
  - **Solution:** Create a conda recipe to set specific versions for all dependencies

We set versions precisely
5- A journey of a team of engineers in learning packaging technology

- **Step 5:** Write a conda recipe for FROGS
  - **Problem:** conda complains when using too specific library version
5- A journey of a team of engineers in learning packaging technology

- **Step 5**: Write a conda recipe for FROGS
  - **Problem**: conda complains when using too specific library version
  - **Solution**: use X.Y instead of X.Y.Z for all library versions
5- A journey of a team of engineers in learning packaging technology

- **Step 6**: run FROGS tests
  - No problem, only one solution: it works!

- **Step 7**: test upload/install using a personal conda channel

- **Step 8**: deployment on bioconda channel => FROGS 2.0.1
5- A journey of a team of engineers in learning packaging technology

- **Step 9:** check all 20 FROGS wrappers using PLANEMO
  - add `<requirement>`, `<stdio>` and `<test>` sections
  - no functional tests available: we made them

- **Step 10:** FROGS wrappers available on TestToolshed
5- A journey of a team of engineers in learning packaging technology

- **Step 11**: deploy that new FROGS package in Galaxy to run a 3 days FROGS training... a success!
6- Results

- FROGS pipeline available as a Conda package
  - https://anaconda.org/bioconda/frogs
  - https://bioconda.github.io/recipes/frogs/README.html

- FROGS wrappers available in Test Toolshed
  - https://testtoolshed.g2.bx.psu.edu/view/oinizan/frogs_2_0_0/e1170d614e52

...and a happy team!
7- Conclusion

- Our two key issues when we started the project:
  - Are these technologies easy to learn? ✔
  - Will we be able to package FROGS in a reasonable amount of time? ✔

⇒ YES!

- Now we gonna publish the package on the toolshed (autumn 2018)
8- Take home messages

- Provide software tools as Conda packages
- Do not let Conda install default dependencies
- Be careful with Conda channels order
- Provide softwares with tests
- Provide separate Conda packages for complex softwares
- Use a logging framework instead of “/dev/null” redirection
Thank You

And thanks to:

- Géraldine Pascal
- Maria Bernard
- The Galaxy community

FROGS dev-team

- Ifb
- INRA
- Ifremer